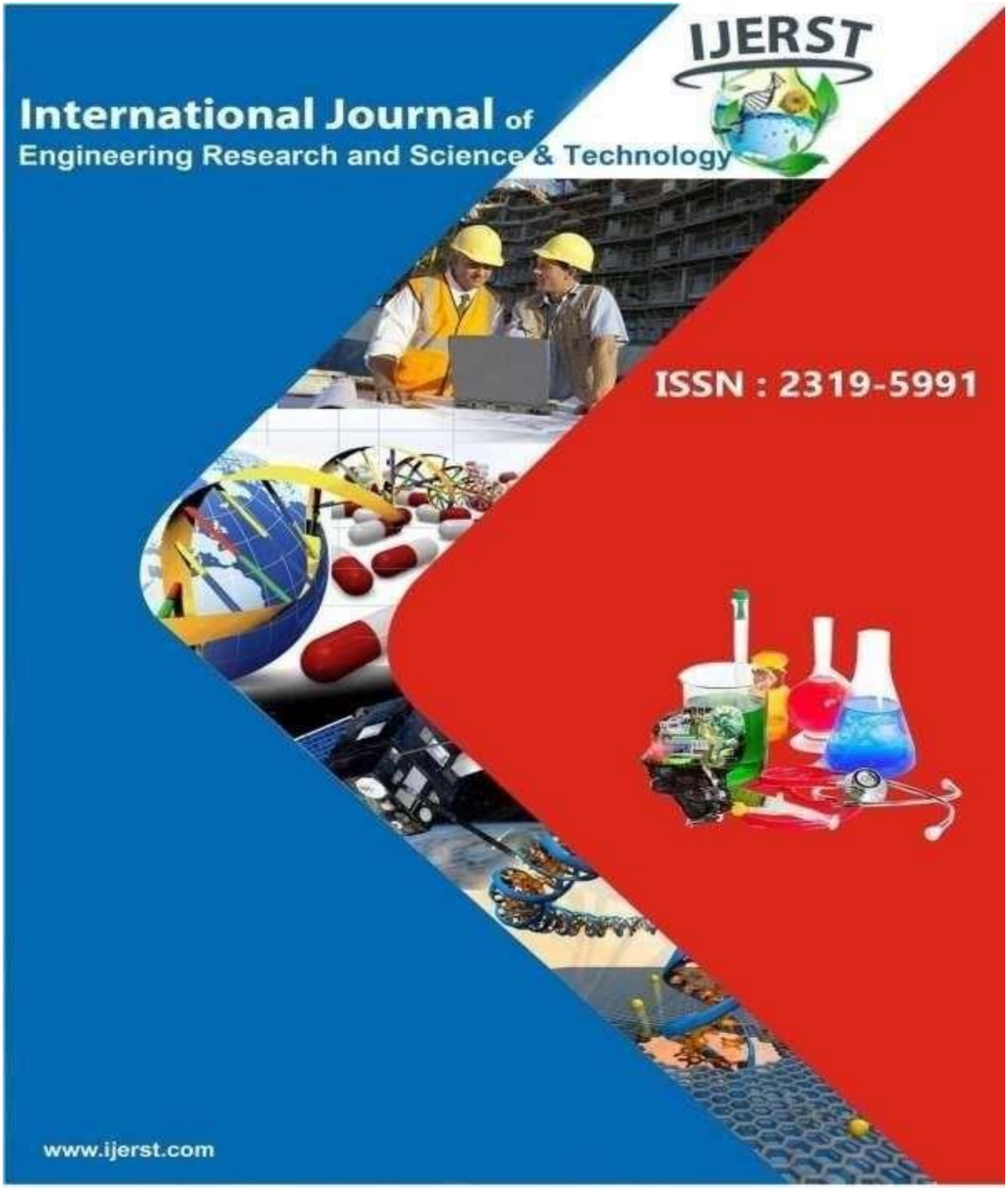


IJERST

International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 1, 2024

DEEP LEARNING BASED COLON POLYP SEGMENTATION WITH TRANSFORMERS

Vuppula Meghana[1], G Praveen Babu[2]
[1]PG Scholar, Department of IT, meghanavuppula26@gmail.com
[2]Assoc Professor, pravbob@jntuh.ac.in
University College Of Engineering Science And Technology, Jawaharlal Nehru Technological University
Hyderabad

ABSTRACT

For computer-aided clinical support systems, automatically analyzing endoscopic pictures presents a challenge when it comes to polyp identification. With encouraging outcomes, models based on VGG19 networks, transformers, and their combinations have been proposed for polyp segmentation. Nevertheless, those methods are either limited to simulating the polyps' local appearance or do not provide multi-level feature representation for spatial dependency throughout the decoding process. To overcome these drawbacks, a novel network called Colon-Former is proposed in this research. An encoder-decoder architecture called Colon Former may simulate semantic information at both the encoder and decoder branches across a long distance. The encoder is a transformer-based lightweight architecture designed to model global semantic relations at several scales. A hierarchical network structure called a decoder is used to enhance feature representation by learning multi-level features. Besides, a refinement module is added with a new skip connection technique to refine the boundary of polyp objects in the global map for accurate segmentation. Extensive experiments have been conducted on five popular benchmark datasets for polyp segmentation, including Kaiser, CVC-Clinic DB, CVC-Colon DB, CVC-T, and ETIS-LA rib. Experimental results show that our Colon Former outperforms other state-of-the-art methods on all benchmark datasets.

1. INTRODUCTION

More than 694,000 people die from colorectal cancer (CRC), one of the most prevalent cancers in the world. The most frequent cause of colorectal cancer (CRC) is colon polyps, specifically high-grade dysplasia adenomas. A long-term study found a 3% decrease in the risk of colon cancer for every 1% increase in the probability of adenoma identification. Therefore, early polyp detection and removal is essential for both the prevention and treatment of cancer. Colonoscopy is therefore recognized as the gold standard for identifying colorectal cancer and colon adenomas. In actuality, overworked healthcare systems—especially those in low-resource environments—may cause endoscopies to take less time and overlook more polyps. A review of the literature indicates that between 20 and 47 percent of colon polyps may be absent during endoscopies. Patients may develop elevated risk factors as a result of this. Consequently, there is a critical need for research into creating computer-aided instruments to help endoscopists during endoscopic treatments. The environment for these kinds of systems has altered due to developments in deep learning and artificial intelligence. In an effort to assist physicians in identifying lesions and reduce the miss detection rate, learning algorithms have been developed to be used in computer-aided diagnostic (CAD) systems for the automatic detection and prediction of polyps. In numerous retrospective studies and diagnoses, deep neural networks have demonstrated significant promise in supporting colon polyp diagnosis. A computerized tomography (CAD) system can assist endoscopists in enhancing lesion detection rates, maximizing techniques for high-risk lesion endoscopy, and expanding clinic capacity without compromising diagnostic quality. Robust polyp segmentation using deep learning has become more accurate and efficient. The most popular method in the past has been convolutional neural networks (CNNs). Long-range dependency modeling is a strong suit for Transformers, a deep learning architecture. Incorporating transformers into deep learning models for colon polyp segmentation is enabling this effort to achieve cutting-edge outcomes.

2. OBJECTIVE

A brief assessment of popular approaches and strategies for polyp segmentation is provided in this area of the project. In order to segment medical images, this study begins with an examination of CNN architectures and their variations, particularly U-Net models. Next, as a promising method to enhance a deep neural network's learning feature representation capacity, look at the attention process. This concludes the investigation of the Vision Transformer and its uses in medical image processing and polyp segmentation. We present Colon Former, a new Transformers-based network that is inspired by these methods for modeling multi-scale and multi-level characteristics. Although a CNN decoder and

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 1, 2024

a transformer encoder are both included in the core architecture of our colon former, our method is distinct from the models previously stated in a few aspects. For learning multi-scale features, the encoder in Colon Former is a lightweight Transformer with a hierarchically structured structure. Using feature maps that are taken from encoder blocks at various scales and subregions, the decoder is a hierarchical pyramid structure that can learn from heterogeneous input. Additionally, a refinement module is suggested in order to increase the accuracy of segmentation on small polyps and hard regions. The goal of this study is to precisely locate and define the boundaries of polyps in colonoscopy pictures automatically. This helps identify colorectal cancer early on, which is a serious health issue. The primary goal is to obtain pertinent characteristics that allow polyps to be distinguished from surrounding tissue. After that, determine the likelihood that each pixel in the picture belongs to a backdrop or polyp. Next, create a binary picture (mask) with background pixels labeled as 0 and pixels corresponding to the polyp labelled as 1.

2.1 PROBLEM STATEMENT

Colon-Former, a unique neural network architecture, is proposed by the research to handle the difficulty of accurate polyp identification and segmentation in endoscopic images. Spatial dependencies are captured successfully by Colon-Former, which combines a hierarchical decoder for multi-level feature representation with a transformer-based lightweight encoder for global semantic relation modeling. Segmentation accuracy is improved by a refinement module that uses inventive skip connections to fine-tune polyp borders in global maps. Colon-Former is superior to current approaches, as shown by extensive trials on benchmark datasets, and thus indicates a considerable development in computer-aided clinical assistance systems for endoscopic imaging.

2.2 Existing System

In order to predict accurate depth maps from a single-color image, the current method presents a novel D-Net model for universal use with multiple encoder backbones.

Strengthening global contextual data and combining them with high resolution features allows D-Net to predict depth maps more accurately. Interestingly, our architecture enables automatic monocular depth estimation during end-to-end training, and it can be used with both convolutional and vision transformer backgrounds.

System ablation investigations further demonstrate that a small encoder backbone combined with the suggested D-Net can provide a very compact, effective, and precise monocular depth estimate.

Existing systems use the D-Net model, which uses unique convolutional blocks to capture just certain polyp properties in deep learning models for polyp segmentation.

Existing System Disadvantages:

- The primary D-Net decoder components are described by the current system.
- No matter whatever backbone is utilized, D-Net is a fully end-to-end trainable model.
- The training and operation of deeper CNN architectures may incur higher computing costs.
- In order to have a more thorough grasp of the properties of polyps, transformers may be able to record the interactions between various elements of the image.
- At the convolutional layers, CNNs mostly pay attention to local features. Since polyps have different looks and require relationships between distant image regions to be taken into account, this is the primary drawback for polyp segmentation.

2.3 Proposed System

- For colon polyp segmentation, this paper suggests a unique deep neural network architecture dubbed Colon Former.
- This model learns an effective multi-scale hierarchical feature representation by utilizing the benefits of both the Transformer and CNN architectures. By relaxing it with a residual link, this model additionally enhances the reverse attention with axial attention.
- The network may gradually adjust the polyp border from a coarse global map that the decoder produces thanks to the refinement module.
- Using widely used benchmark datasets, our system's comprehensive studies demonstrate that Colon Former performs noticeably better than current state-of-the-art models.
- (4)Future research will look into sparse or lightweight self-attention layers as a way to lower the computational complexity.

Proposed System Advantages

- Good performance: VGG19 performs well, particularly when trained on big datasets, as demonstrated by its excellent accuracy in a variety of picture classification tasks.
- Transfer learning: You may use pre-trained weights for feature extraction and obtain good performance even with limited data by

using VGG19 as your basis model for transfer learning.

- Interpretability: The layered convolutional and pooling layers of VGG19's straightforward architecture make it comparatively simple to grasp and comprehend the learnt features.
- Community support: The VGG19 design is extensively researched and used, and a sizable community of practitioners and researchers is working on it. This community offers a wealth of information and assistance for implementing the architecture in Python.

3. RELATED WORKS

We give a quick overview of popular approaches and strategies that have been created for polyp segmentation in this section. Initially, we examine CNN architectures and their variations, particularly UNet models, in the context of medical image segmentation. After that, we Examine how a deep neural network's capacity to learn feature representation can be enhanced by studying the attention mechanism as one promising method. Lastly, the Vision Transformer and its uses for medical image processing and polyp segmentation are examined.

4. METHODOLOGY OF PROJECT MODULES

1 Dataset: We established the method to obtain the input dataset for training and testing purposes in the first module. We have utilized the dataset obtained from Polyp Identification.
There are 4000 photos of polyps in the dataset.

2. Importing the required libraries: To do this, Python will be used. Initially, we will import the required libraries, including pandas, numpy, matplotlib, and tensor flow, as well as keras for creating the primary model and sklearn for dividing training and test data and PIL for converting images to an array of integers.

3 Image retrieval: The pictures and their labels will be retrieved. The pictures should then be resized to ((100, 100)) so that they are all the same size for identification. Next, transform the pictures into a numpy array.

4 Dataset splitting: Divide the dataset into test and train subsets. 20% are test data, while the remaining 80% are train data.

(5) Constructing the model
Architecture: VGG19 has a sequential architecture consisting of five max-pooling layers for down sampling and sixteen convolutional layers, each of which is followed by a Rectified Linear Unit (ReLU) activation function. In addition, it features three fully connected categorization layers at the conclusion. The architecture of VGG19 can be summed up as follows:
Enter the input (224x224x3), 16 layers with 3x3 filters, 1 byte of padding, and ReLU activation are called convolutional layers. Pool Size and Stride: 2 x 2 Pool Size; 5 Layers for Maximum Pooling Fully connected layers: three layers with 4096, 4096, and 1000 units, respectively; the third layer for classification uses soft max activation while the previous two layers use ReLU activation.

Availability of Pre-trained Models Tensor Flow and Keras are two well-known deep learning frameworks that offer VGG19 as a pre-trained model. This model was trained on substantial datasets, notably Image Net. A excellent place to start with transfer learning is with this pre-trained model, which may be refined on certain picture classification tasks using smaller datasets.

6. Use the fit function to apply the model and plot the graphs for accuracy and loss: The model will be compiled and used. There will be two in the batch. The accuracy and loss graphs will then be plotted. Our average training accuracy was 99.3%, while our average validation accuracy was 97.6%.

7. Accuracy on test set: 99.7% of the test set was accurate.

8. Saving the Trained Model: The tested model is first saved as a.h5 file before being transmitted to the production-ready environment once it has been trained. Execution can then take place.

5. DATA FLOW DIAGRAM

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 1, 2024

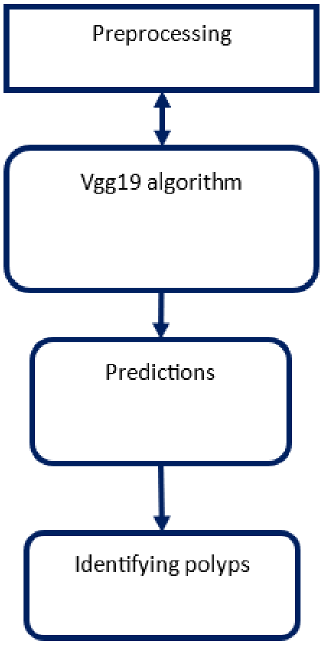


Fig: 7 Flow Diagram

6. ALGORITHM USED IN PROJECT

VGG19 includes a sequential design with 5 max-pooling layers for down sampling, 16 convolutional layers with a Rectified Linear Unit (ReLU) activation function after each, and 3 fully connected layers for classification at the end.

The following succinctly describes the architecture of VGG19:

The input (224x224x3)

Convolutional Layers: 16 layers with ReLU activation, 1 padding, and 3x3 filters

Fully Connected Layers: 3 layers with 4096, 4096, and 1000 units, respectively; ReLU activation in the first two layers and soft max activation in the final layer for classification;

Max-Pooling Layers: 5 layers with a 2x2 pool size and a stride of 2.

7. SYSTEM ARCHITECTURE

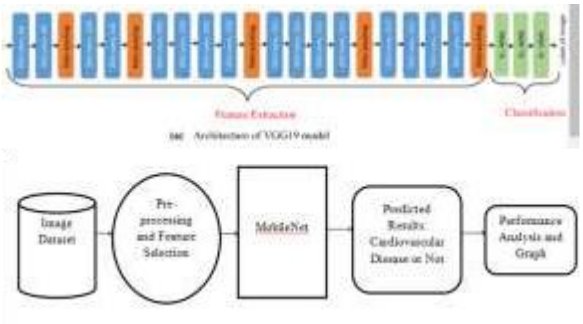
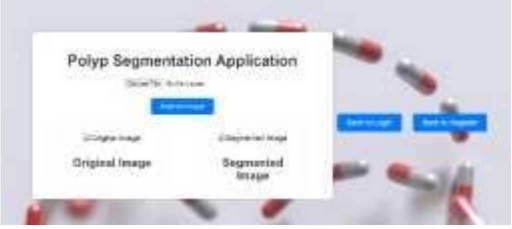
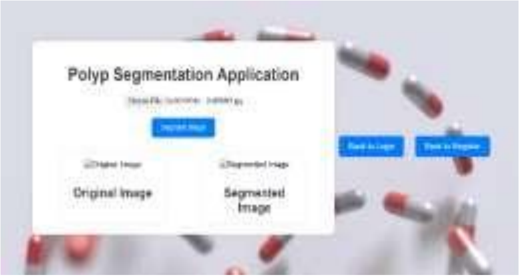

Feature Extraction
Classification
Architecture of VGG19 model
Image Dataset
Pre-processing and Feature Selection
MobileNet
Predicted Results: Cardiovascular Disease or Not
Performance Analysis and Graph

Fig: System Architecture Of Project

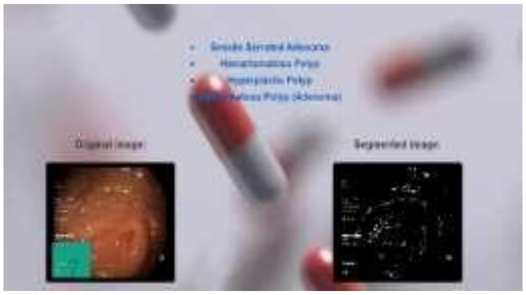
8. RESULTS





ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 1, 2024



9. FUTURE ENHANCEMENT

We plan to explore sparse or lightweight self-attention layers in our upcoming work in order to lower the computational complexity. Furthermore, it is possible to take advantage of alternative architectures that combine CNNs and Transformers.

10. CONCLUSION

The goal of this research is to segment colon polyps using a revolutionary deep neural network architecture called ColonFormer. Utilizing the benefits of both CNN and Transformer architectures, our model acquires a potent multi-scale hierarchical feature representation. By relaxing it with a residual connection, this also improves the reverse attention with axial attention. The network may gradually correct the polyp border from a coarse global map that the decoder produces thanks to the refinement module. Our comprehensive tests demonstrate that, using widely used benchmark datasets, ColonFormer performs noticeably better than current state-of-the-art models. In order to lower the computational cost, we plan to examine lightweight or sparse self-attention layers in future work. Furthermore, different kinds of architectures that combine CNNs and Transformers can also be used.

REFERENCES:

[1] J. Bernal et al., Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge, IEEE Trans. Med. Imag., vol. 36, no. 6, pp. 12311249, Feb. 2017.

[2] D. A. Corley, C. D. Jensen, A. R. Marks, W. K. Zhao, J. K. Lee, C. A. Doubeni, A. G. Zauber, J. de Boer, B. H. Fireman, J. E. Schottinger, V. P. Quinn, N. R. Ghai, T. R. Levin, and C. P. Quesenberry, Adenoma detection rate and risk of colorectal cancer and death, New England J. Med., vol. 370, no. 14, pp. 12981306, 2014.

[3] A. Leufkens, M. G. H. Van Oijen, F. P. Vleggaar, and P. D. Siersema, Factors influencing the miss rate of polyps in a back-to-back colonoscopy study, Endoscopy, vol. 44, no. 5, pp. 470475, 2012.

[4] Mesejo,D.Pizarro,A.Abergel,O.Rouquette,S.Beorchia,L.Poincloux, and A. Bartoli, Computer-aided classification of gastrointestinal lesions in regular colonoscopy, IEEE Trans. Med. Imag., vol. 35, no. 9, pp. 20512063, Sep. 2016.

[5] G. Zhou, X. Liu, T. M. Berzin, J. R. G. Brown, L. Li, C. Zhou, Z. Guo, L. Lei, F. Xiong, Y. Pan, and P. Wang, 951e A real-time automatic deep learning polyp detection system increases polyp and adenoma detection during colonoscopy: A prospective double-blind randomized study, Gastroenterology, vol. 156, no. 6, p. 1511, May 2019.

[6] S. Kudo, Y. Mori, M. Misawa, K. Takeda, T. Kudo, H. Itoh, M. Oda, and K. Mori, Artificial intelligence and colonoscopy: Current status and future perspectives, Digestive Endoscopy, vol. 31, no. 4, pp. 363371, Jul. 2019.

[7] P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H. H.-S. Lu, and V. S. Tseng, Accurate classification of diminutive colorectal polyps using computeraided analysis, Gastroenterology, vol. 154, no. 3, pp. 568575, 2018.

[8] R. Bisschops, J. E. East, C. Hassan, Y. Hazewinkel, M. F. Kaminski, H. Neumann, M. Pellisé, G. Antonelli, M. B. Balen, E. Coron, G. Cortas, M. Iacucci, M. Yuichi, G. Longcroft-Wheaton, S. Mouzyka, N. Pilonis, I. Puig, J. E. van Hooft, and E. Dekker, Advanced imaging for detection and differentiation of colorectal neoplasia: European society of gastrointestinal endoscopy (ESGE) guideline Update 2019, Endoscopy, vol. 51, no. 12, pp. 11551179, 2019.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 111.

[10] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 1001210022.

[11] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 568578.

[12] R. Ranftl, A. Bochkovskiy, and V. Koltun, Vision transformers for dense prediction, in Proc. IEEE/CVFInt.Conf.Comput.Vis.(ICCV),Oct.2021, pp. 1217912188.

[13]S.Zheng,J.Lu,H.Zhao,X.Zhu,Z.Luo,Y.Wang,Y.Fu,J.Feng,T.Xiang, P. H. S. Torr, and L. Zhang, Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers, in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 68816890.

[14] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image

segmentation, in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2015, pp. 234241.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, UNet++: A nested U-Net architecture for medical image segmentation, in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham, Switzerland: Springer, 2018, pp. 311.

[16] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, DoubleU-Net: A deep convolutional neural network for medical image segmentation, in Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS), Jul. 2020, pp. 558564.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, Atrousconvolution, and fully connected CRFs, IEEE Trans. Pattern Anal Mach. Intell., vol. 40, no. 4, pp. 834848, Apr. 2018.

[18] J. Hu, L. Shen, and G. Sun, Squeeze-and-excitation networks, in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 71327141.

[19] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, in Proc. ICLR, Y. Bengio and Y. LeCun, Eds., 2015, pp. 114.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv:1704.04861.

[21] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770778.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, Densely connected convolutional networks, in Proc. IEEE CVPR, Jun. 2017, pp. 47004708.

[23] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, PraNet: Parallel reverse attention network for polyp segmentation, in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2020, pp. 263273.

[24] D. V. Sang, T. Q. Chung, P. N. Lan, D. V. Hang, D. Van Long, and N. T. Thuy, AG-CUResNeSt: A novel method for colon polyp segmen tation, 2021, arXiv:2105.00402.

[25] C.-H. Huang, H.-Y. Wu, and Y.-L. Lin, HarDNet-MSEG: A simple encoder decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS, 2021, arXiv:2101.07172.

[26] P. N. Lan, N. S. An, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and V. D. Sang, NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection, in Proc. Int. Symp. Vis. Comput. Cham, Switzerland: Springer, 2021, pp. 1528.

[27] N. S. An, P. N. Lan, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang, BlazeNeo: Blazing fast polyp segmentation and neoplasm detection, IEEE Access, vol. 10, pp. 4366943684, 2022.

[28] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med. Image Anal., vol. 53, pp. 197207, Apr. 2019.

[29] N. B. Hung, N. T. Duc, T. Van Chien, and D. V. Sang, AG-ResUNet++: An improved encoder decoder based method for polyp segmentation in colonoscopy images, in Proc. Int. Conf. Comput. Commun. Technol. (RIVF), Aug. 2021, pp. 16.

[30] S. Chen, X. Tan, B. Wang, andX.Hu, Reverse attention for salient object detection, in Proc. ECCV, 2018, pp. 234250.